Assignment8

2024-04-10

```
Alex <- read.csv("data_alexithymia.csv", sep = ";")
Alex <- as.data.frame(Alex)

Alex <- Alex[3:24] #remove non-numerical columns
Alex <- Alex[,-c(1)]# remove age as a factor
#train = sample(1:nrow(x), nrow(x)/2)
sample <- sample.int(n = nrow(Alex), size = floor(.75*nrow(Alex)), replace = F)
train <- Alex[sample, ]
test  <- Alex[-sample, ]



test = as.data.frame(test)
train = as.data.frame(train)
#x.test = x[test, ]
#y.test = y[test]
str(Alex)
```

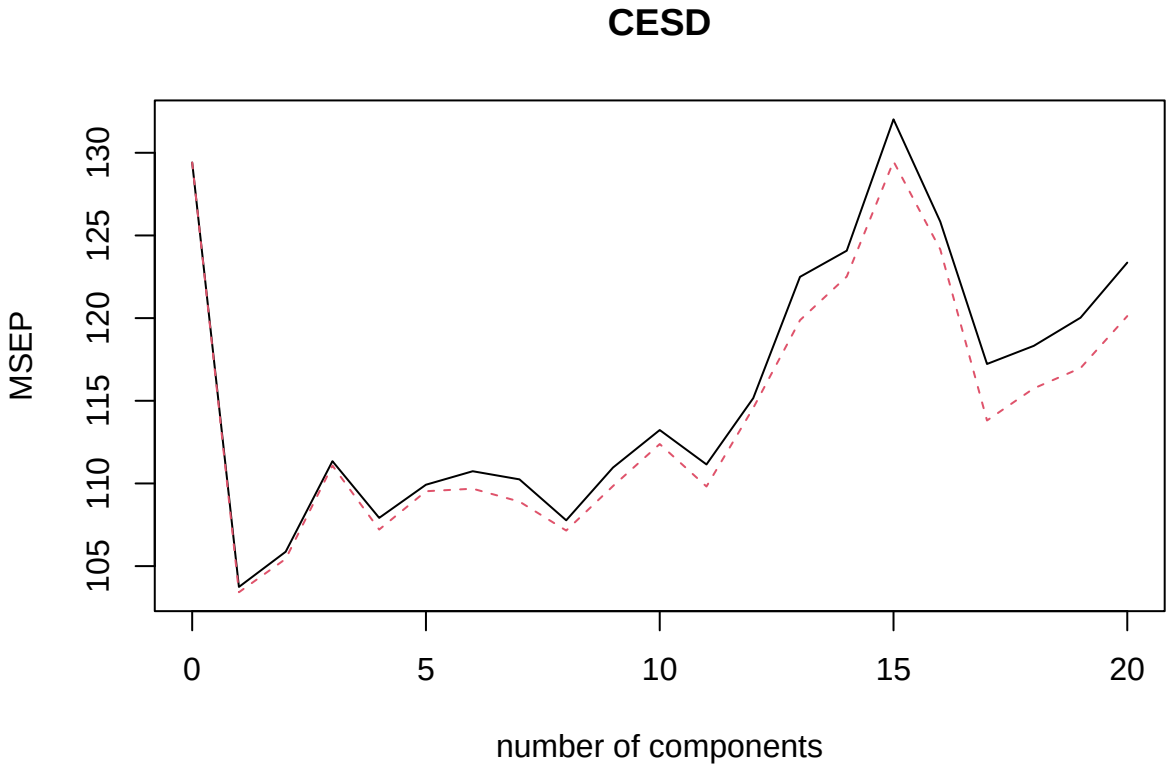
```
## 'data.frame':    122 obs. of  21 variables:
##  $ X.confused         : int  0 3 3 2 1 2 4 2 2 1 ...
##  $ X.right.words      : int  0 3 1 0 2 4 3 2 1 0 ...
##  $ X.sensations       : int  0 0 0 0 0 0 3 4 0 3 ...
##  $ X.describe         : int  3 3 3 4 4 0 2 2 3 4 ...
##  $ X.analyze.problems : int  1 4 3 4 4 3 0 2 2 3 ...
##  $ X.upset            : int  0 2 0 0 1 3 2 2 2 0 ...
##  $ X.puzzled          : int  0 1 0 2 0 2 2 3 2 0 ...
##  $ X.let.happen       : int  2 1 0 0 1 0 1 1 2 1 ...
##  $ X.identify         : int  0 2 1 0 1 3 1 3 1 0 ...
##  $ X.essential        : int  2 3 3 4 4 3 3 2 3 4 ...
##  $ X.feel.about.people: int  1 2 0 0 1 4 4 3 1 0 ...
##  $ X.describe.more    : int  0 0 0 0 2 1 0 2 2 0 ...
##  $ X.going.on         : int  0 1 0 0 0 3 3 4 1 0 ...
##  $ X.why.angry        : int  0 1 3 0 0 2 0 1 1 0 ...
##  $ X.daily.activities : int  2 1 1 4 0 2 1 1 1 0 ...
##  $ X.entertainment    : int  3 1 0 0 0 3 3 2 3 0 ...
##  $ X.reveal.feelings  : int  0 1 0 4 1 4 4 3 3 0 ...
##  $ X.close            : int  4 2 3 4 4 4 2 3 3 4 ...
##  $ X.useful           : int  2 3 2 4 3 3 4 3 3 4 ...
##  $ X.hidden.meanings  : int  2 1 0 4 4 3 1 2 0 1 ...
##  $ CESD               : int  0 23 46 11 8 18 26 16 13 3 ...
```

Principle Component Regression (PCR)

You can also embed plots, for example:

```
## Data:    X dimension: 91 20
##  Y dimension: 91 1
## Fit method: svdpc
## Number of components considered: 20
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV           11.38    10.19    10.29    10.55    10.39    10.48    10.52
## adjCV        11.38    10.17    10.27    10.54    10.35    10.47    10.47
##        7 comps  8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## CV       10.50    10.38    10.53     10.64     10.54     10.73     11.07
## adjCV    10.44    10.35    10.48     10.60     10.48     10.70     10.95
##        14 comps  15 comps  16 comps  17 comps  18 comps  19 comps  20 comps
## CV        11.14     11.49     11.22     10.83     10.88     10.96     11.11
## adjCV     11.07     11.38     11.14     10.67     10.76     10.82     10.96
##
## TRAINING: % variance explained
##       1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
## X       24.94    38.01    46.38    53.79    59.89    65.72    70.58    74.76
## CESD    21.86    22.37    22.37    25.44    26.31    27.71    28.38    28.76
##       9 comps  10 comps  11 comps  12 comps  13 comps  14 comps  15 comps
## X       78.50     81.68     84.44     86.96     89.41     91.50     93.47
## CESD    30.06     30.52     32.81     32.92     36.17     36.17     38.06
##       16 comps  17 comps  18 comps  19 comps  20 comps
## X        95.27     96.85     98.23     99.21    100.00
## CESD     38.88     42.29     42.70     44.81     44.94
```

```
validationplot(pcr_fit_train, val.type="MSEP" )
```




```
pcr_predict = predict(pcr_fit_train, test[, 1:20], ncomp = 20)
pcr_mse = mean((pcr_predict -test$CESD)^2)
pcr_mse
```

```
## [1] 77.12181
```

Partial Least Squares (PLS) Regression

```
set.seed(123123123)


pls_fit_train = plsr(CESD ~ ., data = train, scale=TRUE, validation = "CV")

summary(pcr_fit_train)
```

```
## Data:    X dimension: 91 20
##  Y dimension: 91 1
## Fit method: svdpc
## Number of components considered: 20
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##        (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV           11.38    10.19    10.29    10.55    10.39    10.48    10.52
## adjCV        11.38    10.17    10.27    10.54    10.35    10.47    10.47
##        7 comps  8 comps  9 comps  10 comps  11 comps  12 comps  13 comps
## CV       10.50    10.38    10.53     10.64     10.54     10.73     11.07
## adjCV    10.44    10.35    10.48     10.60     10.48     10.70     10.95
##        14 comps  15 comps  16 comps  17 comps  18 comps  19 comps  20 comps
## CV        11.14     11.49     11.22     10.83     10.88     10.96     11.11
```

```
## adjCV     11.07     11.38     11.14     10.67     10.76     10.82     10.96
##
## TRAINING: % variance explained
##       1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
## X       24.94    38.01    46.38    53.79    59.89    65.72    70.58    74.76
## CESD    21.86    22.37    22.37    25.44    26.31    27.71    28.38    28.76
##       9 comps  10 comps  11 comps  12 comps  13 comps  14 comps  15 comps
## X       78.50     81.68     84.44     86.96     89.41     91.50     93.47
## CESD    30.06     30.52     32.81     32.92     36.17     36.17     38.06
##       16 comps  17 comps  18 comps  19 comps  20 comps
## X        95.27     96.85     98.23     99.21    100.00
## CESD     38.88     42.29     42.70     44.81     44.94
```

```
validationplot(pls_fit_train, val.type="MSEP" )
```